DOCUMENT RESUME

ED 358 527 EA 024 958

AUTHOR Cooley, William W.
TITLE The Difficulty of the Educational Task.
INSTITUTION Pittsburgh Univ., Pa. Learning Research and Development Center.
PUB DATE 12 May 93
NOTE 17p.
PUB TYPE Reports - Research/Technical (143) -- Viewpoints (Opinion/Position Papers, Essays, etc.) (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; *Census Figures; Educational Diagnosis; *Educational Testing; Elementary Secondary Education; *Evaluation Methods; *Evaluation Problems; Public Schools; *Student Characteristics
IDENTIFIERS Census 1990; *Pennsylvania

ABSTRACT

Comparison of student test scores between states, school districts, and even schools continues to be a popular measure of student achievement. However, these comparisons reveal little about the quality or effectiveness of educational programs, only the varying difficulty of educating different populations of students. This report uses U.S. Census data and information on Pennsylvania school districts to explain differences in the difficulty of the educational task. Demographic data on adult high school graduation rates, single-parent homes, and poverty levels can indicate the difficulty of educating students. Data indicate that as the percentage of children living in poverty in single-parent households with parents who did not graduate from high school increases, academic performance decreases. This analysis indicates that all three census variables affect student performance, and if any one were eliminated, performance could be expected to increase. Similarly, National Assessment of Educational Progress data from 42 states, when compared with census data, showed correlations between demographic characteristics and student performance. Changes should be made in how student test scores are compared and interpreted to provide a truer picture of student performance. (JPT)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Pennsylvania Educational Policy Studies

PEPS is a joint effort of the University of Pittsburgh's School of Education and the Learning Research & Development Center
This is policy paper number 16 in this series

ED358527

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
K. Malarz
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

The Difficulty of the Educational Task

by

William W. Cooley

University of Pittsburgh

May 12, 1993

The purpose of this series of papers is to contribute to a more informed debate about critical policy issues facing Pennsylvania's public schools. This PEPS series draws upon a data base that has been established here at the University of Pittsburgh under the direction of William Cooley in cooperation with the Pennsylvania Department of Education.

Reactions can be shared:

by mail: LRDC, Pgh., PA 15260
by PittVAX: COOLEY
by FAX: 412-624-9149
by PENN*LINK: PEPS
by phone: 412-624-7085
by INTERNET: COOLEY@VMS.CIS.PITT.EDU
by chat: room 743, LRDC

EA 024 958

BEST COPY AVAILABLE

The Difficulty of the Educational Task

William W. Cooley
Pennsylvania Educational Policy Studies
University of Pittsburgh

Comparisons of student test score results have become a national pastime. The National Assessment of Educational Progress (NAEP) is now comparing states, the states are comparing school districts or schools within their state, and people (politicians, reporters, parents, educators, taxpayers, etc.) continue to make invalid inferences about the relative effectiveness of the educational systems being compared. A big problem is that such comparisons reveal little or nothing about differences in the quality or effectiveness of the educational programs that are represented by those results. What those differences in test results primarily reveal are differences in the difficulty of the educational task, which is a function of the differences in the populations being served by those different systems.

The purpose of this PEPS report is to show how the latest U.S. Census data can be useful in developing indicators of how states and school districts differ in the difficulty of their educational task, and how and why those demographic differences explain the differences in national and state test score results. The implications of this for national and state wide testing programs, as well as implications for equity in school district funding in Pennsylvania, are also considered.

Indicators of the Difficulty of the Educational Task

The PEPS project has recently completed the merger of U.S. Census data with our extensive state data base. This required our being able to combine census counts for Pennsylvania's 2,584 Minor Civil Divisions (MCD's are townships, boroughs, cities, etc.) into the 500 operating school districts.[1] For example, the census provides estimates of the number of persons in each MCD that are not high school graduates. If a school district serves six MCD's, their results are combined into an estimate of the percent of that school district's population who have not graduated from high school. In the average Pennsylvania school district, 25 percent of the persons age 18 and over did not complete high school (as shown in Table 1). But this percentage varies dramatically across the state. In one district, only 4 percent of the adult population did not complete high school, while in another district over half of the adults did not. A district with a well-educated adult population has an easier educational task than a district that does not.

Similarly, school districts vary in the percentage of their students that are being raised in poverty. In the average district, about 13% of the school age children are being raised in families that are below the poverty level, as defined in the 1990 census. But here again the variation among the 500 districts is very large, with some districts

TABLE 1

Census Indicators of the Difficulty of Task
(for 500 PA school districts)

	Mean	Range		Correlations		
		Lowest	Highest	S.P.	N.H.S.	Pov.
% Single Parent	17	6	58	1.00	.31	.66
% Not H.S. Grads	25	4	52	.31	1.00	.57
% Poverty	13	0	53	.66	.57	1.00

having no poverty children, while in other districts over half of the children come from poor homes. Districts with lots of poverty children have a more difficult educational task than do districts with very few students from low income families.

Another census-derived indicator of the difficulty of the educational task is the frequency of single parent homes. There is only one parent in about 17% of the families with school age children for the average school district, but in some districts that percentage is greater than 57%, while in others it is less than 6%. Although there is some controversy about the possible negative effects of this factor on student achievement, districts in which there are many single parent families

seem to have a more difficult educational task than do districts with very few such families.

Table 1 summarizes the descriptive statistics for these three census derived variables, as well as the degree to which they correlate with each other.[2] The correlations indicate that districts with high poverty tend to also have fewer high school graduates (.57) and more single parent families (.66), but that the relationship between percent high school grads and single parent families is not nearly as strong (.31) as the other two relationships.

Difficulty of Task and Student Achievement

It is very important to emphasize that in using these census derived indicators of the difficulty of the educational task we are describing school districts, and not individual children. Certainly not all poor children, nor children whose parents did not graduate from high school, nor children with only one parent, have difficulty learning in school. The point here is that as the percentage of such children increases in a school district, the lower will be the average performance of all children in the district on a common test administered to all districts. Let us now turn to the validity of that claim.

From the PEPS database it is possible to estimate the differences in student performance among the 500 operating school districts in Pennsylvania, using results from the Test of Essential Learning and

Literacy Skills (TELLS), which was last given to Pennsylvania's third, fifth and eighth graders in 1991. The test samples what students should be expected to know and be able to do in reading and mathematics by those grade levels. A district composite was created that reliably describes the differences among these districts in reading and mathematics achievement, based upon those test results.[3]

This student performance composite, when correlated with the 1990 census data that are descriptive of the difficulty of the educational task for each of the 500 school districts in Pennsylvania, yields a multiple correlation of .78. This means that over 60 percent of the variation in the average student performance among these school districts can be explained by those three simple census factors, leaving only about 40 percent to be explained by all other possible factors, including other demographic variables besides these three. In other words, comparing districts on such a state-wide test reveals more about the difficulty of their educational task than about the quality of their educational program.

The results from this analysis also indicate that all three census variables make a unique contribution to the prediction of student performance. In other words, if any of these three census predictors were dropped there would be a significant loss in the predictive power of the resulting multiple regression equation. In some districts

performance is explained more by poverty while in others, for example, adult educational level may be what is contributing more to the prediction. But all three are useful predictors, even in combination with the others.

The National Assessment of Educational Progress

The 1990 census results also make it possible to determine how well the difficulty of the educational task variables can explain the state comparisons for the NAEP mathematics results. In 1992 the NAEP mathematics test was administered in a manner which allowed comparisons among 42 states (including the District of Columbia). The publication of these results became a major media event. The NAEP reports usually include a chart which they claim "provides a sound basis for making appropriate comparisons in average proficiency across states and territories because it shows whether or not the average between pairs of states is statistically significant". That is, the observed state differences were probably not due to sampling error or measurement error. However, the reports do not make clear what an "appropriate comparison" might be, given its statistical significance.

Table 2 reports the results of deriving the three difficulty-of-task indicators for the 42 states for which NAEP mathematics average proficiency estimates were available. For example, in the average state, 24 percent of the population age 18 and over did not graduate from high

TABLE 2

Census Indicators of the Difficulty of Task (for 42 states in NAEP)

	Range			Correlations		
	Mean	Lowest	Highest	S.P.	N.H.S.	Pov.
% Single Parent	22	16	53	1.00	.33	.47
% Not H.S. Grads	24	16	34	.33	1.00	.80
% Poverty	17	7	32	.47	.80	1.00

school, but in one state (Colorado) only 16 percent did not, while in another (Mississippi) 34 percent did not graduate from high school. The other two census variables are also summarized in Table 2. That very high percentage of single parent families (53 percent) in Table 2 was for the District of Columbia.

In terms of the pattern of correlations, the strongest relationship was between percent of school-age children from poverty homes and the percent of the adult population in the state that did not graduate from high school (.80). States with lots of poverty tended to have fewer high school graduates, as one would expect. The weakest relationship was between percent single parent families with school age children and percent of adults that did not graduate from high school (.33).

TABLE 3

Correlation of Difficulty and Performance

Census Indicator of Difficulty	PA TELLS (500 districts)	NAEP Math (42 states)
% Single Parent Families	-.63	-.73
% Not H.S. Grads	-.62	-.71
% Poverty Children	-.66	-.72

Table 3 shows how each of the three census variables correlates with the NAEP math means for those states. Even though their inter-correlations varies from .80 to .33 as shown in Table 2, the three census variables have very similar relationships to student math performance. The negative correlations indicate that high math means tend to be associated with low percentages on these census indicators, with each of them explaining about half of the variance in state math means. Table 3 also shows how the three census indicators correlated with the TELLS district means.

When these three state census indicators were combined in a multiple regression for predicting NAEP mathematics means for these 42 states, a multiple correlation of .89 was found. This indicates a very strong relationship between these family variables and student

performance on this test. In other words, if you rank order states on the basis of the difficulty of their educational task, you get about the same rank orders as are produced using the NAEP average proficiency for these states. Therefore one can clearly not make inferences about the relative quality of the math programs in these 42 states when over 75 percent of the variation in the math means among these states can be explained by the nature of the populations being served by the schools in those states.

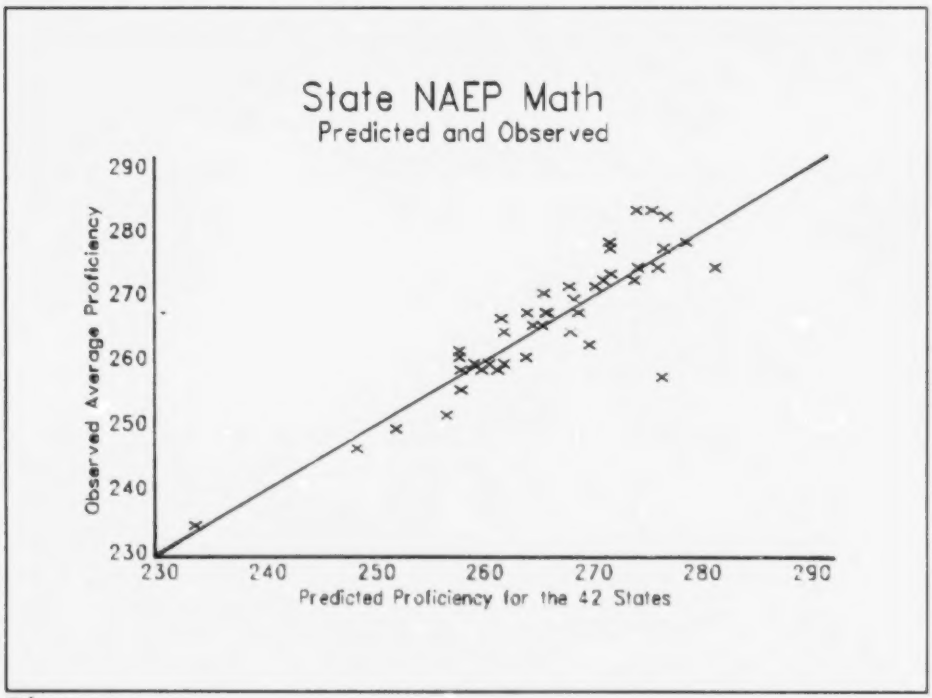

Figure 1

Figure 1 might help to clarify this point. The horizontal axis in that figure represents the combination of the three census variables that best

predicts the NAEP math results. The vertical axis represents the average proficiency for each of these 42 states. The diagonal line represents what the predicated values would be if the prediction were perfect. Most states (X's) lie very close to the prediction line, indicating that student performance was about what would be expected given this census information. In those states above the diagonal line, the students are doing better than expected, and in those states below the line the students are performing lower than would be expected from these census data.

One reason why test results are more dependent upon home differences than differences in the quality of the educational program is that tests such as TELLS or NAEP are not keyed to a specific curriculum. They are not examinations on what has just been taught. The test questions represent a sample of what tends to be taught at the particular grade level for which the test is designed, but the test is not necessarily a good fit to any particular school's curriculum. Such tests are very sensitive to family differences.

There are at least two things to be done that would make comparisons of student achievement differences more valid. One would be to have tests that clearly reflect a common curriculum for all educational systems being compared. This is now happening in some states, and that is an encouraging trend. The other is to statistically

adjust the observed test means in a manner that takes into account the differences in the populations being compared.

One of the reasons given for not adjusting test results for home differences is the problem of seeming to encourage low expectations for systems with the more difficult educational task. For example, "The students in that system did not do well on the test, but what do you expect given the kinds of students they have to educate."

Another reason for not adjusting test scores is that our conceptual models (and their research basis) for determining what to adjust for may be inadequately specified. For example, a state's location well below the prediction line in Figure 1 may not reflect an inadequate math program. What it may reflect is the fact that an important demographic variable has been left out. It is also possible that one or more of the demographic variables used in the predication has only a spurious, non-functioning relationship to average student proficiency.

Sometimes it does make sense to report the observed, unadjusted test results. If, for example, the test questions reflect a desired and accepted standard for student performance, then unadjusted scores make it possible to see how well those students are meeting the standard. Similarly, if such tests were scaled so that they are comparable over time, then unadjusted scores make it possible to see whether those students are making progress toward those standards.

NAEP is making some progress on both of those fronts, but much remains to be done.

A very large problem with unadjusted scores is that there are educational systems that are doing a good job with difficult-to-educate students, but their successes go unrecognized and unrewarded when their unadjusted results are unfavorably compared with systems that had the easy job. Teacher frustration is a frequent by-product of the practice of releasing and comparing unadjusted test results.

When comparisons are being made to support arguments about the relative effectiveness of education systems (Japan vs. the United States, or Colorado vs. Mississippi, or Upper Merion vs. Chester Upland), it is essential that student test results be adjusted for relevant differences in the populations being served by those systems. Not doing so results in invalid inferences. We need to establish what those most relevant population differences are. This demonstration of the power of linking census data to student performance information illustrates why it is important to do so.

Implications for School Finance Reform

The fact that the systems with the most difficult educational task tend to have the fewest resources available for improving their educational systems is another reason for bringing these census data into the discussion of test score results. In Pennsylvania, as in most

states, expenditures per pupil varies as a function of the local tax base available to support the local school district. This in turn is highly negatively related to the effort (resources) required to educate their students. That is, districts with the easiest educational task tend to have the most to spend, and districts with the most difficult task tend to have the least to spend.

For example, the multiple correlation between district expenditures per pupil and these three census variables is .53, showing that districts with the easiest educational task have the most to spend. The percent non-high school graduates in the district population was the most negatively related to expenditures per pupil. It is possible that a more highly educated population demands a higher quality (or at least more expensive) educational program, but the data are more consistent with the fact that districts with the more difficult educational task tend to have a smaller per capita tax base, and thus fewer resources for supporting a stronger educational program.

Policy Implications

Test score averages for nations, states, districts or schools should not be released without helping people make valid inferences as to what the results indicate and do not indicate.

Test scores cannot be compared for systems that do not share a common curriculum. If people insist upon a state wide test, they must first be willing to agree upon a common curriculum framework or the desired learning outcomes for that state, and be willing and able to include adjustments for the populations being compared.

More valid inferences can be made if systems are being compared to an established standard than if systems are compared to each other. Performance indicators that are comparable over time are much more useful in guiding system improvement than when they not comparable over time. Most state testing programs have this weakness.

School finance reform must be a part of any state's effort at systemic educational reform, since districts with the most difficult educational task tend to have fewer educational resources than do districts with the easier educational task.

Endnotes

1. The PEPS project is very indebted to John Senier of the PA Department of Education for his assistance in relating census data to school districts. His MCD and school district "crosswalk" included how data for an MCD that is served by two different school districts could be proportioned.

2. These three census indicators were derived from Summary Tape File 3 on five CD-ROM disks which were recently released by the US Bureau of the Census. The Pennsylvania data were from STF3A (MCD summary level), and the state level census data for the NAEP analysis were from STF3C. The NAEP state average proficiency results are available from just about every newspaper in the country.

3. The most reliable measure of the student achievement differences among these 500 districts is the principle component of the six means (reading and math for grades 3, 5 and 8) available for each district. The six factor loadings ranged from .82 to .90, with 76% of the variance explained by this one factor.